



2Market Sales and Customer Data Analysis

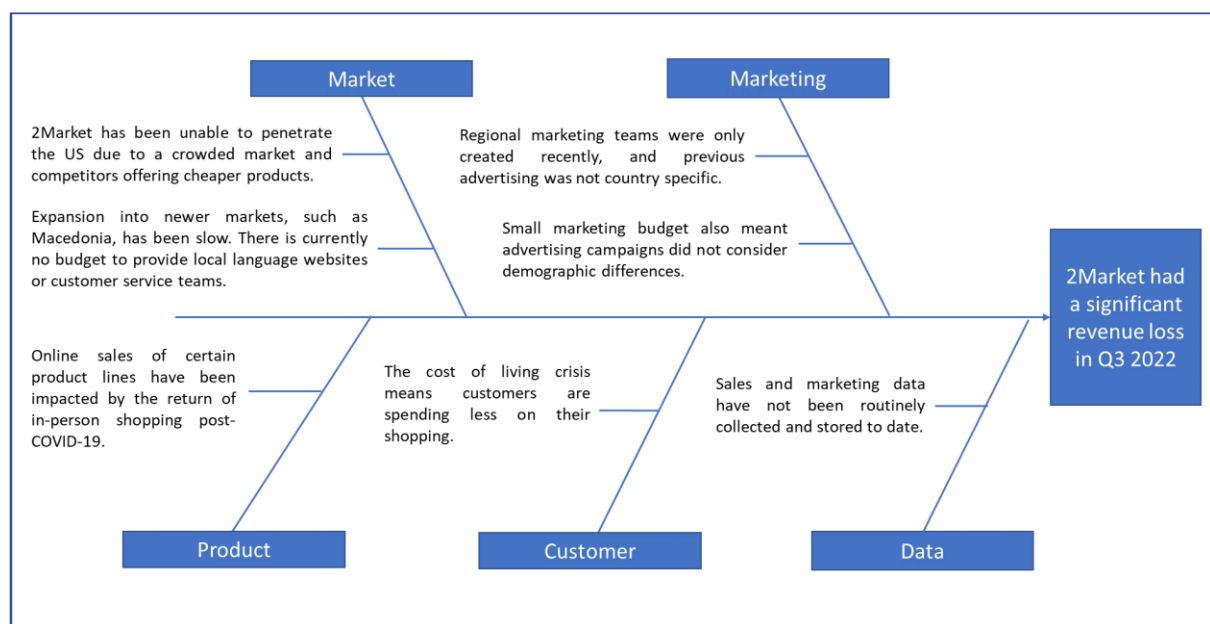
2Market Data Analytics Project – December 2022

1. Background

2Market commissioned this project to better understand its customers, top products and markets by sales, and most effective advertising channels.

A kick-off meeting with stakeholders revealed a loss in Q3 2022 revenue that requires immediate cost savings and increased sales. Further investigation shows this was mainly caused by US competition, ineffective marketing, unsuccessful market expansion, and decreased online sales of certain products post-COVID.

Figure 1. Fishbone diagram showing the causes of 2Markets Q3 2022 revenue loss



This project aims to support senior management make decisions about where to focus the business and help regional marketing teams improve campaigns and identify the best advertising channels by product. Stakeholders acknowledged there was no data linking ad conversion to sales but agreed an estimation was sufficient.

2. Analytical Approach

The Marketing_data and Ad_data files were viewed as text to see delimiters and check for any unusual data.

Data cleaning:

- A spellcheck and search for blank cells was conducted – no issues detected.
- Conditional Formatting was used on the ID column ('Unique customer ID') to check for duplicate rows – none were found.
- In the Marital_Status column, 'YOLO,' 'Alone,' and 'Absurd' were replaced with 'Null,' as it could not be determined if these meant Single, Divorced, or Widow.
- In the Education column, '2n Cycle' was changed to 'Master.'
- Data in the Country column was changed to full country names, while several column titles were changed to be more informative.
- Outliers were removed for Age and Income. The upper age limit was 89 so customers aged 121, 122 and 128 years were removed, while eight with incomes above \$118,351 were removed.
- In the Dt_Customer column, Text to Columns was used to fix any dates not in date format.
- The data type for each column was checked and changed accordingly. For any column with a monetary value, the \$ symbol was replaced with a blank and then changed to number format.

Data analysis:

Initial exploration with Excel revealed a strong relationship between increasing customer income and sales for all product types.

- In an interim meeting with the regional marketing teams, income was selected as the most important customer demographic for initial campaigns.

Figure 2: Scatterplot showing the relationship between customer income and total spend



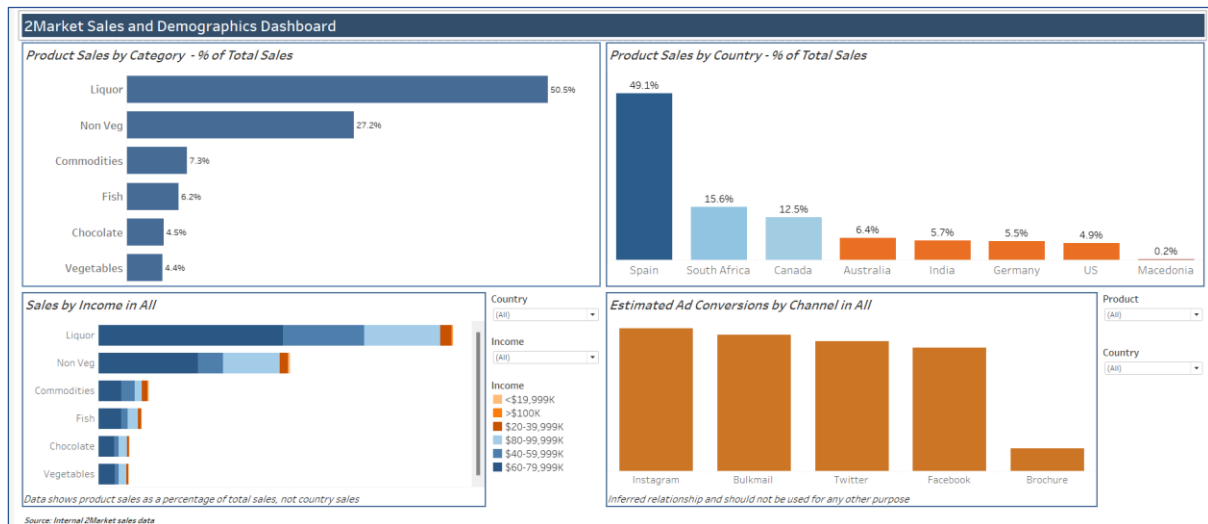
SQL was used to infer a relationship between product sales and successful lead conversions by advertising channel.

- To do this, a new table was created which joined the Marketing_data and Ad-data using an inner join and customer ID ('ID') as the primary key. The new table contained the country, total sales per product, and lead conversion data for each customer.
- A query was created to show the sum of all successful conversions for each advertising channel when there were sales of each product type (using > 0) by country. Product type was added to a new 'product' column, and the query was repeated for each product using a UNION operator (*see Appendix for syntax*).
- Results showed the inferred number of successful ad conversions for each advertising channel by product type and country.
- Results were saved as the Ad_conversion_by_product CSV file to be uploaded onto Tableau for visualisation.

3. Dashboard Design and Development

The 2Market Dashboard consists of four visualizations and has been designed without red or green colours as one member of the regional marketing team is colour blind. Tooltips have been added to all charts to ensure they are interpreted accurately, whereas labels have only been used in static charts to avoid overcrowding in the dashboard. Chart titles have been edited to be meaningful and dynamic titles also added where relevant.

Figure 3: The 2Market Dashboard



The **Product Sales by Category** and **Product Sales by Country** visualisations were created from the Marketing_data and are intended to quickly show overall sales performance for senior management. This will allow them to make decisions on which markets to focus on or withdraw from, and which products to focus on or discontinue.

- **Product Sales by Category** shows product sales as a percentage of total sales. This was created with a calculated field ('Total Sales'), with further calculated fields created for each product as a percentage of 'Total Sales.' Aliases were created on charts e.g., '% Sales Liquor' changed to 'Liquor.'
- The **Product Sales by Country** chart shows sales by country as a percentage of total sales. A map was originally considered, but a bar chart was deemed more impactful.

The **Sales by Income** visualisation was created from the Marketing_data and is intended to support the regional marketing teams assess the impact of customer income on product sales.

- The data has been broken down by adding income categories to the colour marks, with income categories created using bins of \$20,000.
- The chart includes both a filter and dynamic title for Country so data can be seen at a regional level.
- A caption has been added to note that tooltips show data as a percentage of total sales and not sales by country.

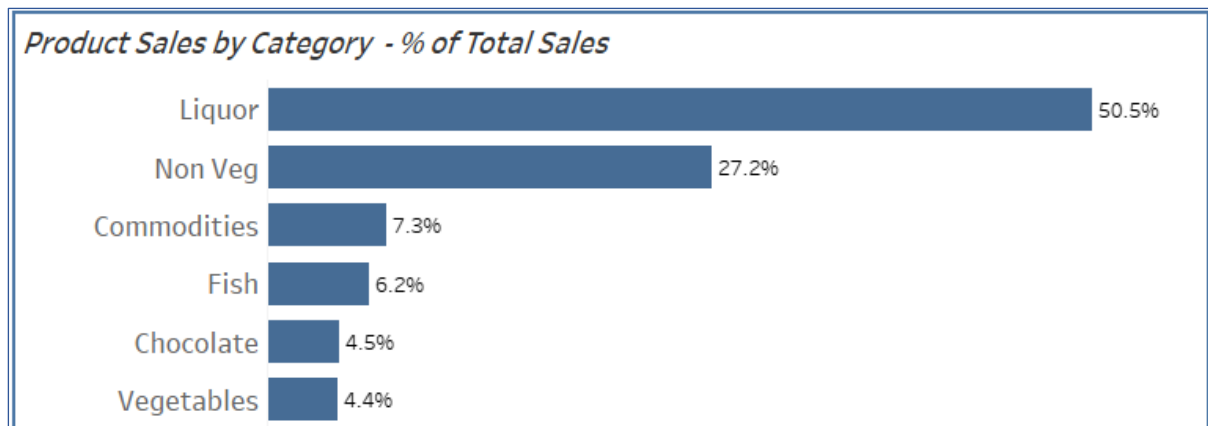
The ***Estimated Ad Conversions by Channel*** visualisation is also intended for the regional marketing teams to see the inferred relationship between successful ad conversions and product sales by country.

- This was created from the Ad_conversion_by_product data from SQL.
- A dynamic title has been added to show data by country.
- Labels have not been included as the chart is only meant to show which channel may be best.
- A caption highlights that the data infers a relationship and should not be used for other purposes.

4. Patterns, Trends, and Insights

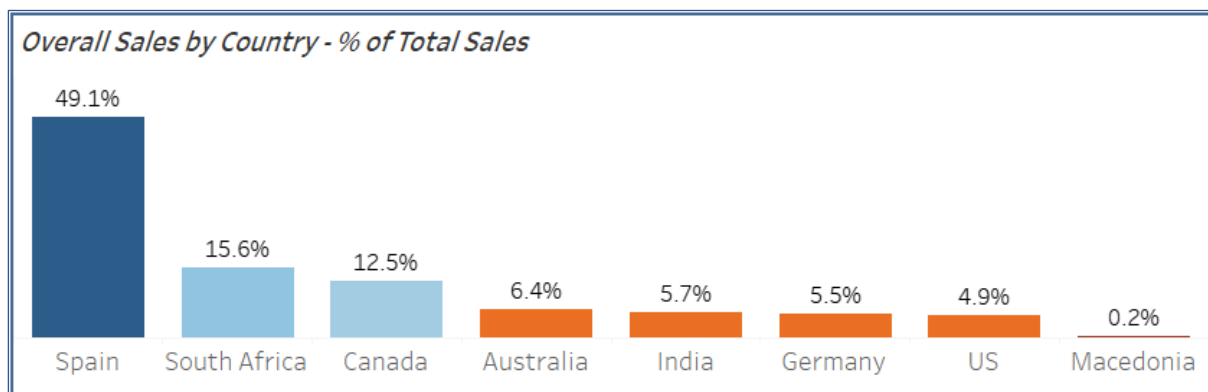
Analysis shows that liquor and non-vegetables account for >75% of sales, while **vegetables and chocolate make up <10% of sales**. This should enable management to decide whether to discontinue certain products in order to reduce storage and distribution costs.

Figure 4: Total sales by product



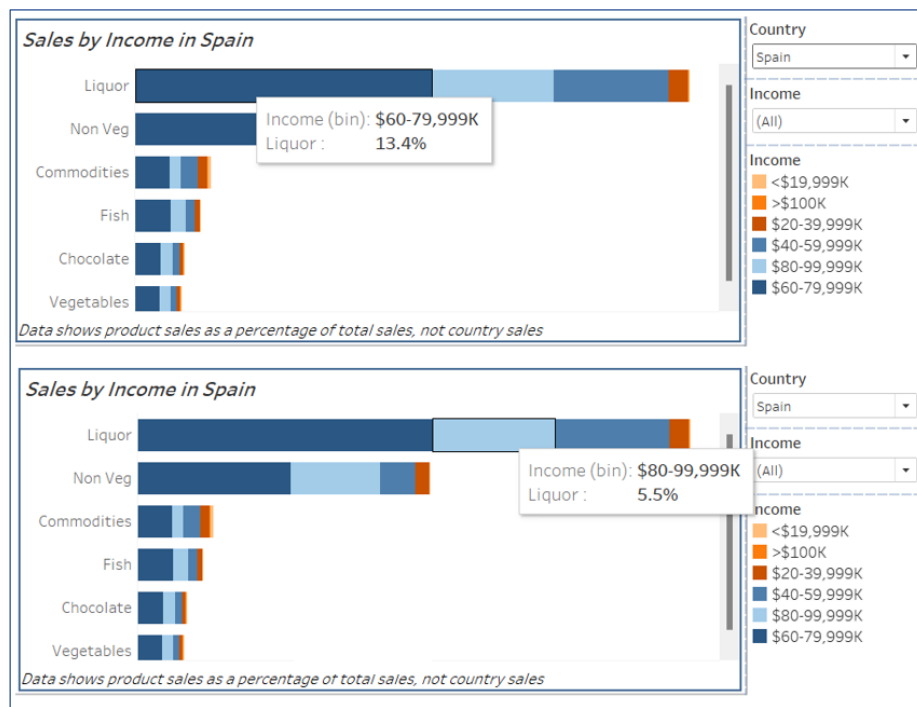
Data shows Spain as the leading market, with almost 50% of sales. The **US makes up <5% while Macedonia has ~0.2%**. This should allow management to decide whether to withdraw from certain countries in order to make operational savings.

Figure 5: Total sales by country



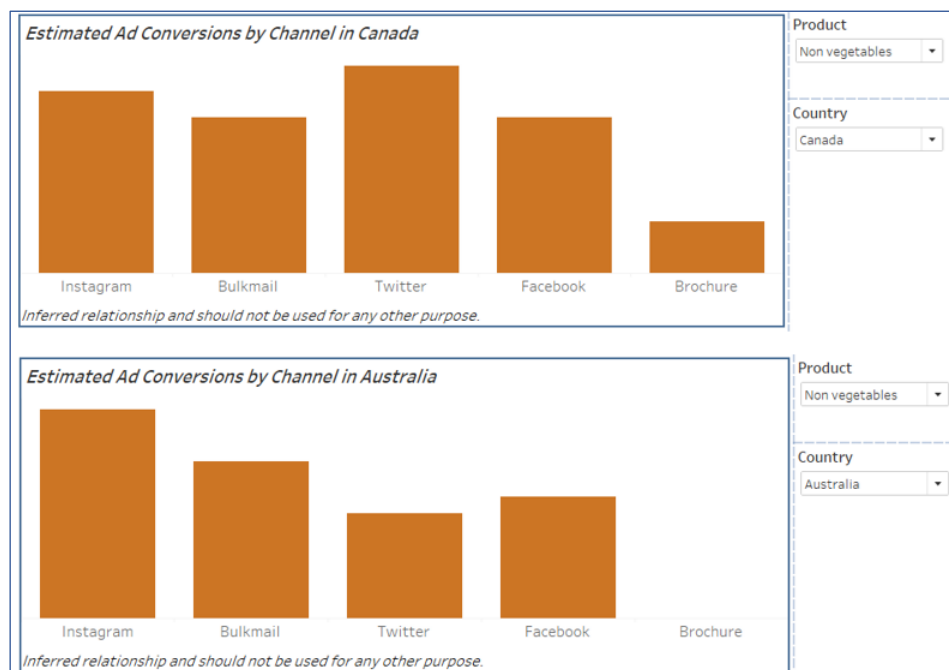
Data also shows that **different customer income impacts total sales**. For example, customers in Spain earning \$60-\$79,999 account for 13.4% of total liquor sales, whereas those earning \$80-\$99,999 account for 5.5%. This will support regional marketing teams in targeting different customer groups with personalised campaigns, offers, or discounts.

Figure 6: Data showing the impact of income on products sales



Data shows the **relationship between ad conversions and sales is impacted by country**. For example, sales of non-vegetables are driven by Twitter in Canada but by Instagram in Australia. This will support the regional marketing teams choose effective ad channels.

Figure 7: Data showing the impact of different advertising channels on products sales by country



Recommendations for further exploration:

- Gather data to see true sales performance over time by product and country, including a breakdown of online and in-person sales.
- Gather data on the impact of campaigns on sales, such as campaign date, country, and product.
- Conduct further analysis on the impact of other demographics on sales, such as age.

5. Appendix

1. Create new table

```
CREATE TABLE ad_conversion AS
```

```
SELECT md.id as customer_id, md.marital_status, md.liquor, md.vegetables, md.non_veg,  
md.fish_products,
```

```
md.chocolates, md.commodities, md.country, ad.twitter_ad, ad.instagram_ad, ad.facebook_ad,  
ad.bulkmail_ad, ad.brochure_ad
```

```
FROM public.marketing_data md
```

```
JOIN public.ad_data ad
```

```
USING (id);
```

2. Query relationship between product sales and effective advertising conversions by country

```
SELECT adc.country, 'liquor' product, SUM(adc.twitter_ad) as Twitter, SUM(adc.instagram_ad) as  
Instagram,
```

```
SUM(adc.facebook_ad) as Facebook, SUM(adc.bulkmail_ad) as Bulkmail, SUM(adc.brochure_ad) as  
Brochure
```

```
FROM public.ad_conversion adc
```

```
WHERE adc.liquor > 0
```

```
GROUP BY adc.country
```

```
UNION
```

```
SELECT adc.country, 'vegetables' product, SUM(adc.twitter_ad) as Twitter, SUM(adc.instagram_ad)  
as Instagram,
```

```
SUM(adc.facebook_ad) as Facebook, SUM(adc.bulkmail_ad) as Bulkmail, SUM(adc.brochure_ad) as  
Brochure
```

```
FROM public.ad_conversion adc
```

```
WHERE adc.vegetables > 0
```

```
GROUP BY adc.country
```

```
UNION
```

```
SELECT adc.country, 'non_veg' product, SUM(adc.twitter_ad) as Twitter, SUM(adc.instagram_ad) as  
Instagram,
```

```
SUM(adc.facebook_ad) as Facebook, SUM(adc.bulkmail_ad) as Bulkmail, SUM(adc.brochure_ad) as  
Brochure
```

```
FROM public.ad_conversion adc
```

WHERE adc.non_veg > 0

GROUP BY adc.country

UNION

SELECT adc.country, 'fish_products' product, SUM(adc.twitter_ad) as Twitter,

SUM(adc.instagram_ad) as Instagram,

SUM(adc.facebook_ad) as Facebook, SUM(adc.bulkmail_ad) as Bulkmail, SUM(adc.brochure_ad) as Brochure

FROM public.ad_conversion adc

WHERE adc.fish_products > 0

GROUP BY adc.country

UNION

SELECT adc.country, 'chocolates' product, SUM(adc.twitter_ad) as Twitter, SUM(adc.instagram_ad) as Instagram,

SUM(adc.facebook_ad) as Facebook, SUM(adc.bulkmail_ad) as Bulkmail, SUM(adc.brochure_ad) as Brochure

FROM public.ad_conversion adc

WHERE adc.chocolates > 0

GROUP BY adc.country

UNION

SELECT adc.country, 'commodities' product, SUM(adc.twitter_ad) as Twitter,

SUM(adc.instagram_ad) as Instagram,

SUM(adc.facebook_ad) as Facebook, SUM(adc.bulkmail_ad) as Bulkmail, SUM(adc.brochure_ad) as Brochure

FROM public.ad_conversion adc

WHERE adc.commodities > 0

GROUP BY adc.country

ORDER BY country, product;